

Introduction to meta-analysis in psychology

Marco Marelli

Spirit of the class

- A general idea on what meta-analysis is and how it works
- Hands-on experience on how to run a meta-analysis (in R)
- Starting point rather than a comprehensive presentation of the method

Meta-analysis

- Statistical method aimed at combining evidence from different studies (on a **common topic**) to provide a single answer to a **specific research question**
- A meta-analysis is considered a **secondary source**:
 - A study that is founded on evidence originally reported somewhere else

Meta-analysis: the general idea

- Every individual study is **limited** (sample size, power, various biases...)
- However, the assumption is that there is a **unique truth** behind all the studies on a given topic
- In order to understand this truth, results from different studies can be pooled together and aggregated
- From statistical testing to assessing **the magnitude of the effect**

Meta-analysis in a nutshell

- An «**analysis of analysis**»
- From evaluating, in a number of different studies, the evidence about a given phenomenon as well as its reliability...
- ...to aggregating the different results in some form of **weighted synthesis** that (allegedly) capture the truth behind all the studies

Meta-analysis in a nutshell

- Given an effect investigated by a number of studies, a meta-analysis can tell us:
 - The **mean and variance** of the effect at the population level
 - The **variability** of the effect across studies
 - The potential impact of **moderator variables**

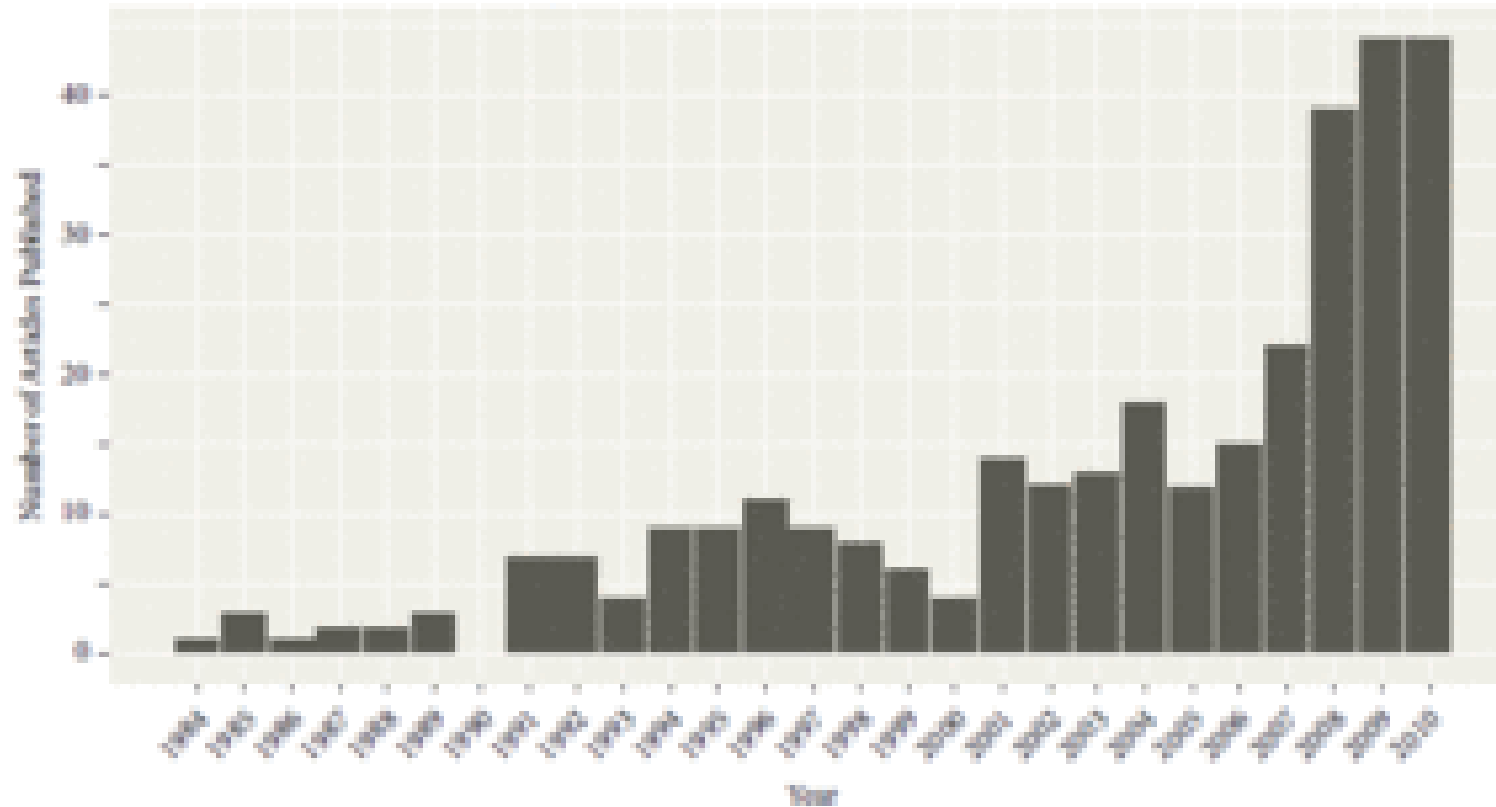
A bit of history

- Discussions about the possibility of combining different studies together (Fisher, 1938; Pearson, 1938)
- Fisher (1944): “When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance”
- Cochran (1953) discusses a method of averaging means across independent studies (and concepts like inverse variance weighting and homogeneity testing are introduced)

A bit of history

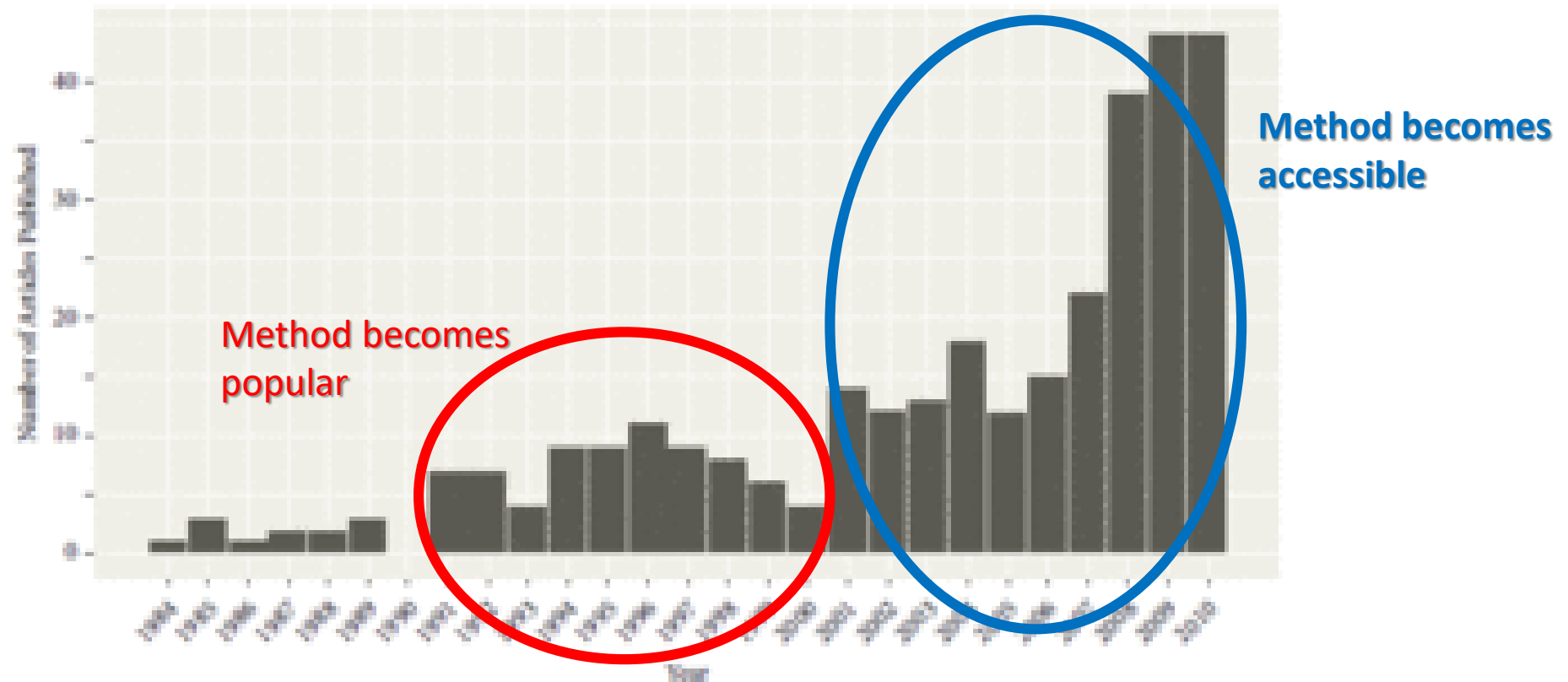
- In 1977, Smith and Glass statistically aggregates the effects from 375 studies in clinical psychology (in psychotherapy outcome)
- The term “meta-analysis” is coined (Glass, 1976)
- Rosenthal (1978): theoretical perspective on meta-analysis
- Books are published on the topic:
 - Rosenthal (1984), Hedges and Olkin (1985), Hunter and Schmidt (1990)

How is it going?



Meta-analysis papers in clinical psychology from 1984 to 2010

How is it going?



Meta-analysis paper in clinical psychology from 1984 to 2010

Advantages

- A well-defined quantitative approach to combining empirical findings
- Non-hierarchical approach, not favoring earlier or later studies
- Protects against over-interpreting differences across studies
- Can handle a large numbers of studies (vis-à-vis literature reviews)
- Guide the research progresses and the development of new studies

Difficulties

- Not effortless!
- Comparability of the studies is in the “eye of the beholder”
- Less objective than it may seem:
 - Which studies to include?
 - Which study attributes to consider and how?
- Analysis of between-study differences is fundamentally correlational
- Retrieving data is not always possible

How to run a meta-analysis (Field 2013)

1. Exploring the literature
2. Establishing inclusion criteria
3. Exporting effect sizes
4. Running the actual meta-analysis
5. (Running some follow-up/advanced analyses)
6. Writing the paper!

1- Exploring the literature

- On the basis of the research question under investigation
- Aimed at identifying the relevant empirical research
- Main instruments: repositories, search engines, databases
- Published research is easy to identify but...

1- Exploring the literature

- File drawer problem!
- Any meta-analysis should account for the publication bias: significant results are 8 times more likely to be published than non-significant results!
- Other than typical scientific papers, it is important to consider:
 - Results from conference talks and posters
 - Results from master/doctoral dissertations
 - Unpublished data from different labs (contact experts in the field!)

2- Establishing inclusion criteria

- Not every study should be included in the actual analysis
- Potential issues need to be avoided:
 - Including poorly-conducted studies
 - Comparing «apples and oranges»

Although meta-analysis might seem to solve the problem of variance in study quality because these differences will “come out in the wash,” even one red sock (bad study) amongst the white clothes (good studies) can ruin the laundry.

(Field & Gillette, 2010)

2- Establishing inclusion criteria

- Specific and detailed criteria must be established in order to choose which studies are to be included in the analysis
 - Amount and type of participants
 - Amount and type of items
 - Specific manipulations
 - Covariates considered
 - Confounding variables controlled for
 - Instrument used

3- Exporting effect sizes

- For each included study, we need a way to quantitative define the reported effect
- In some cases we are lucky and we have a set of studies with highly homogenous dependent measures
- That's not always the case!!
- Often we have to apply some transformation to the data

3- Exporting effect sizes

- Standardize the effects within each study so that these can be combined and compared.
- We can apply standard measures of effect size:
 - Cohen's d for group differences
 - Pearson's r for association between variables
- Sometimes it may be difficult:
 - Some papers do not report effect sizes
 - Some papers lack enough details to compute them ourselves
- Over and above the effect size, we need some associated dispersion measure (e.g., the standard error)

4- Running the actual meta-analysis

- Aim: to estimate the effect in the population (and a confidence interval) by combining the effect sizes from different studies using a weighted mean.
- Weights?
 - Values reflecting the sampling precision of each effect size
 - Associated with sample size (larger samples provide more reliable data)

4- Running the actual meta-analysis

- Fixed-effects or Random-effects?

4- Running the actual meta-analysis

- Fixed-effects model:
 - Sampled from a population where effect size is fixed
 - Sample effect sizes should be homogeneous
 - Based on comprehensive set of representative studies
 - The model characterizes the scope of the present research
 - Error derived from sampling studies from a population of studies
 - Valid inferences only for the studies included in the analysis

4- Running the actual meta-analysis

- Random-effects model:
 - The average effect size in the population varies randomly
 - Effect sizes should be heterogeneous
 - Populations are sample from a “superpopulation”
 - Afford inferences about a broader population
 - Have also error due to sampling from a superpopulation
 - Inferences that generalize beyond the studies in the meta-analysis

4- Running the actual meta-analysis

- When in doubt, **go for the random-effects model** (Field, 2013)
 1. typically, social science data have heterogeneous effect sizes
 2. typically, in psychology we want to generalize our inferences beyond studies included in the meta-analysis
 3. wrongly applying a random-effects model does not have dramatic consequences
- Different methods can be applied (for details, see Field, 2005)

5- Running some follow-up/advanced analyses

- How to estimate **publication bias**?
- Fail-safe N (Rosenthal, 1979)
 - estimating the number of unpublished studies needed to turn a significant population effect size into a non-significant one
- Funnel plot (Light & Pillerner, 1984)
 - plotting effect sizes against standard errors (or sample size, variance, etc)
- Trim&Fill (Duval & Tweedie, 2000)
 - funnel plot is trimmed and missing studies from the truncated part are estimated. Next, k artificial studies are added (“filled”) to the negative side of the funnel plot

5- Running some follow-up/advanced analyses

- Any **moderator**?
- If there is variability in the effect sizes, what is the reason?
- Essentially running a regression model with effect sizes as DV
- Potential moderators could be either continuous or categorical

6- Writing the paper!

- It is important to make it clear why the meta-analysis is run
- There are papers dedicated to a single meta-analysis
- There are papers where meta-analysis is used to (and even requested to) establish a case study

6- Writing the paper!

- Meta analyses used to be applied to evaluate whether an effect, reported across several studies, is really significant
 - combine effect sizes from comparable studies
- Meta analyses are also used (currently) to examine relationships between study properties and the effect sizes they obtain
 - combine effect sizes from as many studies as possible, carefully coding for study properties
 - moderator analysis becomes crucial

A side effect from word
recognition – an example

What makes the following word-sets different?

A	B
Widow	Whisk
Cheer	Cheek
Herb	Helm
Poet	Pond
Quiet	Quest
Train	Trail

What makes the following word-sets different?*

Transparent Set	Opaque Set
Widow	Whisk
Cheer	Cheek
Herb	Helm
Poet	Pond
Quiet	Quest
Train	Trail

**These words are often found as targets for different conditions in the majority of morphological priming studies.*

A classic masked priming paradigm

transparent

opaque

• TARGET

SPEAK

ARCH



PRIME

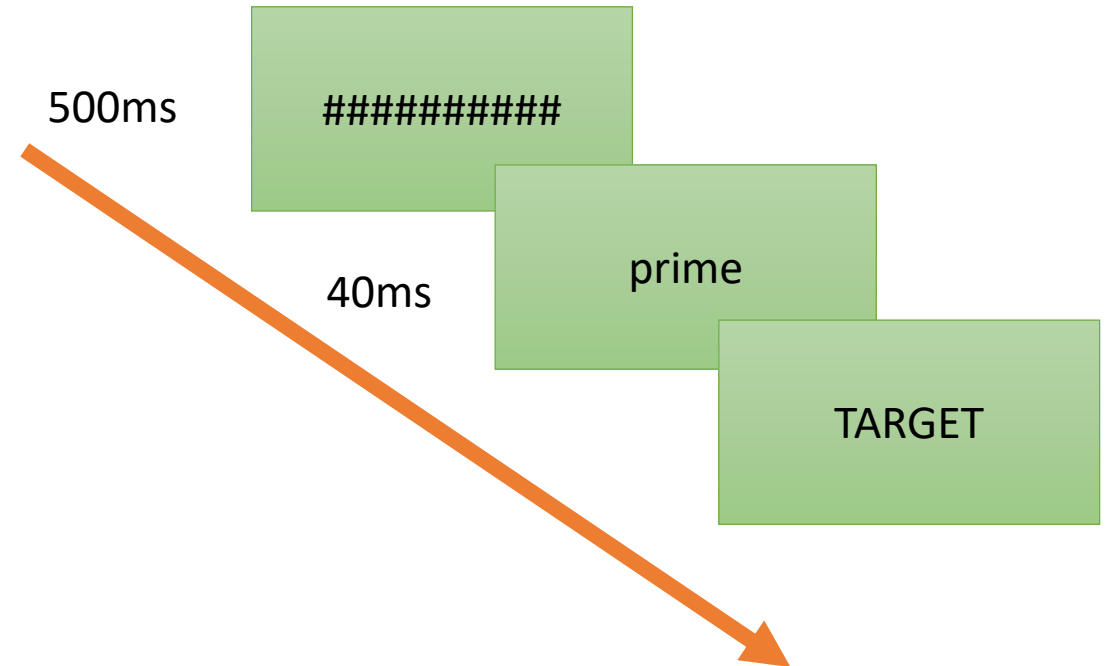
SPEAKER

ARCHER

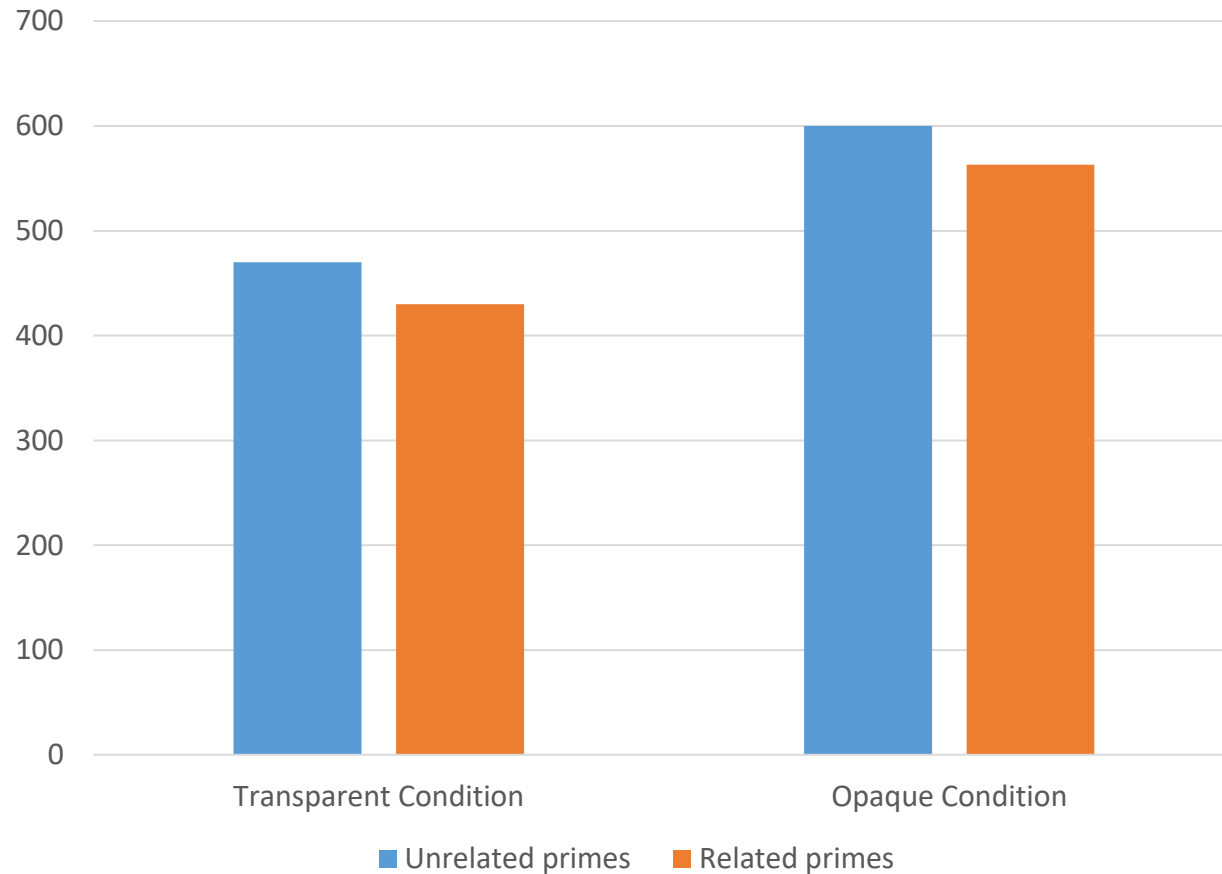
UNREL. PRIME

POTTERY

POCKET



A “stem transparency” effect?



There is a priming effect for both conditions

However, also a main effect of “transparency” seems to be there (and usually ignored)

Targets in the transparent condition have shorter RTs than targets in the opaque condition, irrespective of the prime

Is that really the case?

A “stem transparency” effect?

- Let’s test the intuition with a meta-analysis
- R package: metafor
- To run the model: `rma()`
- Forest plot: `forest()`
- Funnel plot: `funnel()`

A “stem transparency” effect?

- Inclusion criteria:

1. have stems as visually-presented target words
2. employ lexical decision as task
3. are run on native speakers of the language of interest
4. include both a transparent and an opaque condition
5. adopt a between-target design

(as it is often the case, we were not able to retrieve data for all the studies meeting the criteria above)

A “stem transparency” effect?

Study	Effect	S.E.	Language
Rastle et al. (2000) Exp1	31 ms	7.55	English
Devlin et al. (2004)	34 ms	15.02	English
Rastle et al. (2004)	30 ms	8.27	English
Diependaele et al. (2005) Exp2	15 ms	6.51	French
Kazanina et al. (2008)	19 ms	8.16	Russian
Marslen-Wilson et al. (2008) Exp1	17 ms	4.98	English
Diependaele et al. (2011) Exp1	3 ms	5.99	English
Kazanina (2011) Exp2	13 ms	8.93	Russian
Andrews et al. (2013)	-3 ms	9.42	English
Marelli et al. (2013)	37 ms	6.50	Italian